Journal club & Exercise

Jinho
Postdoctoral Fellow
Linding lab, DTU

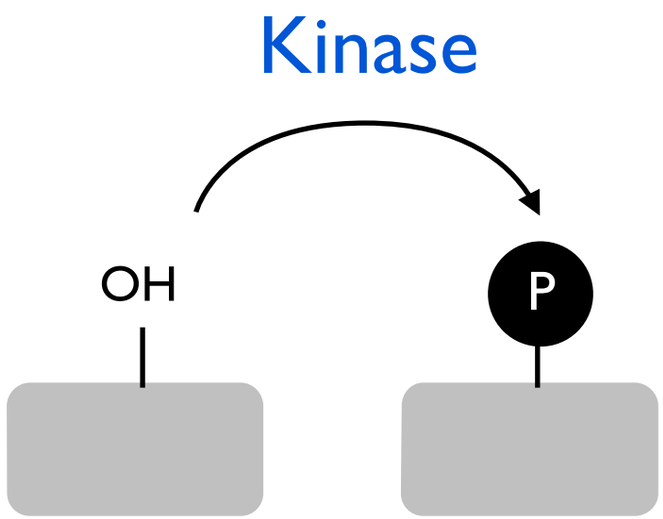
- http://wiki.bio.dtu.dk/teaching/index.php/IntroduktionTilSystemBiologi2014

Paper 1: NetPhorest (Miller et al.,Linear Motif Atlas for Phosphorylation-Dependent Signaling. Science Signaling, 2 September 2008, Vol 1, Issue 35, p. ra2)
Paper 2: NetworKIN (Linding et al., Systematic discovery of in vivo phosphorylation networks. Cell. 2007 Jun 29;129(7):1415-26)

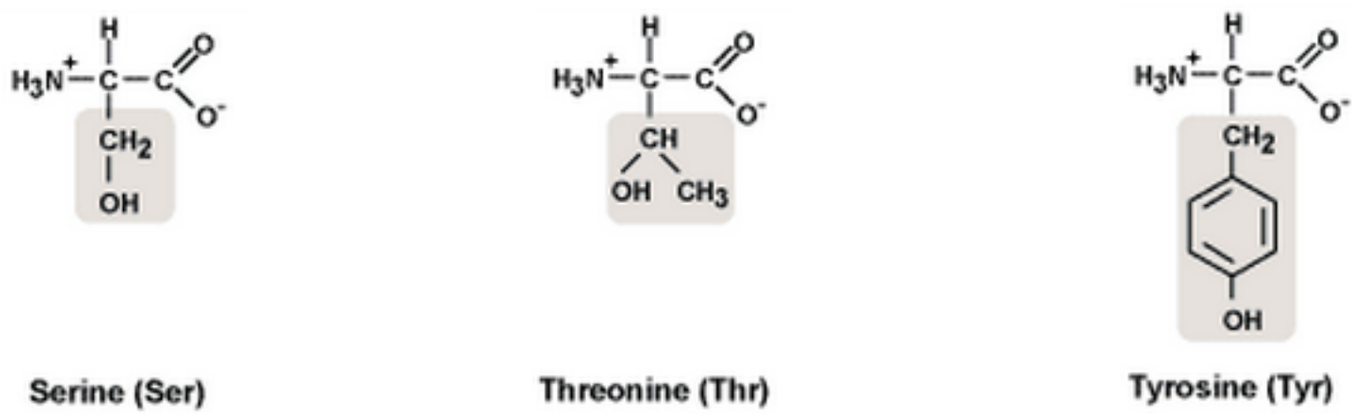
KinomeXplorer (http://kinomexplorer.info)

Protein kinase

Protein kinase is an enzyme that puts a phosphate on its target proteins (substrates).



Phosphorylatible amino acids

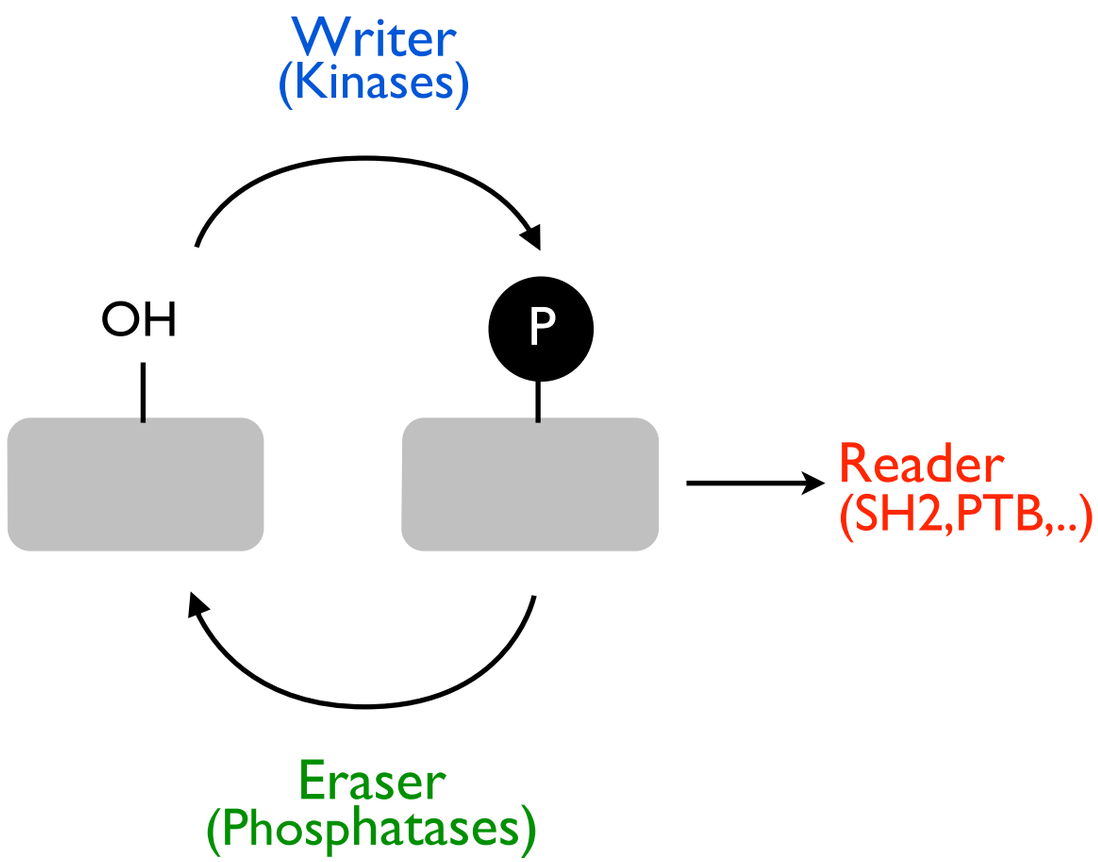


S T Y

http://www.personal.psu.edu

Kinases in cellular signaling process

Writer-Reader-Eraser toolkit



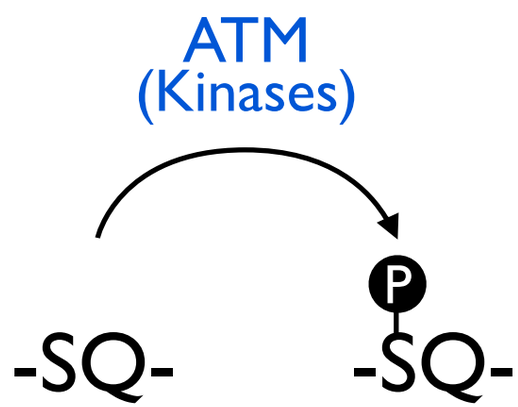
Switch

www.examiner.com

Motif (word in natural language)

Kinases recognize specific sequence motifs and phosphorylate them.
(Every kinase has its favorite words.)

eg)



Different kinases recognize different motifs. Some of them allow several amino acids at one position, which can be represented by position specific score matrix (PSSM).

eg)

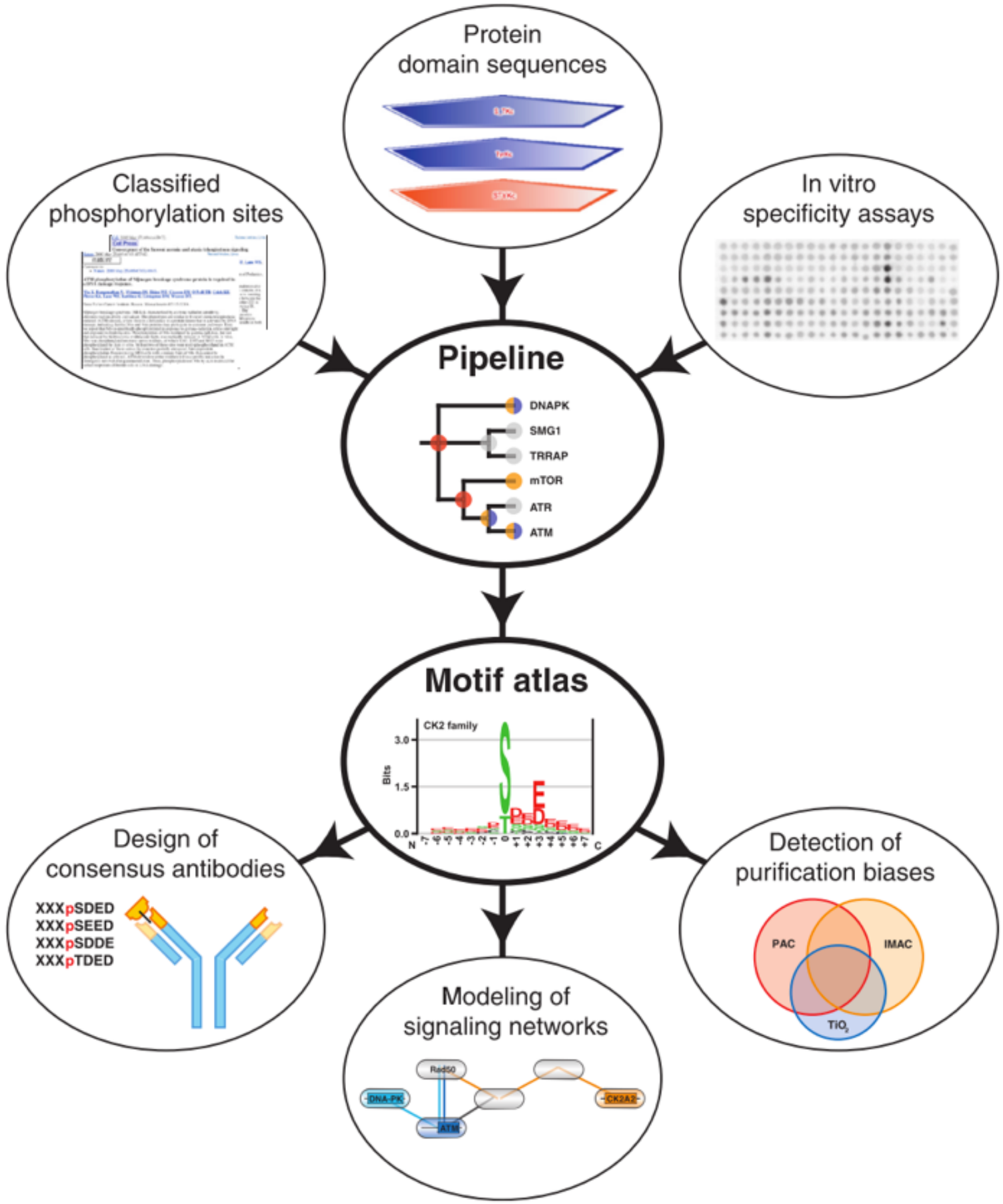
NetPhorest pipeline

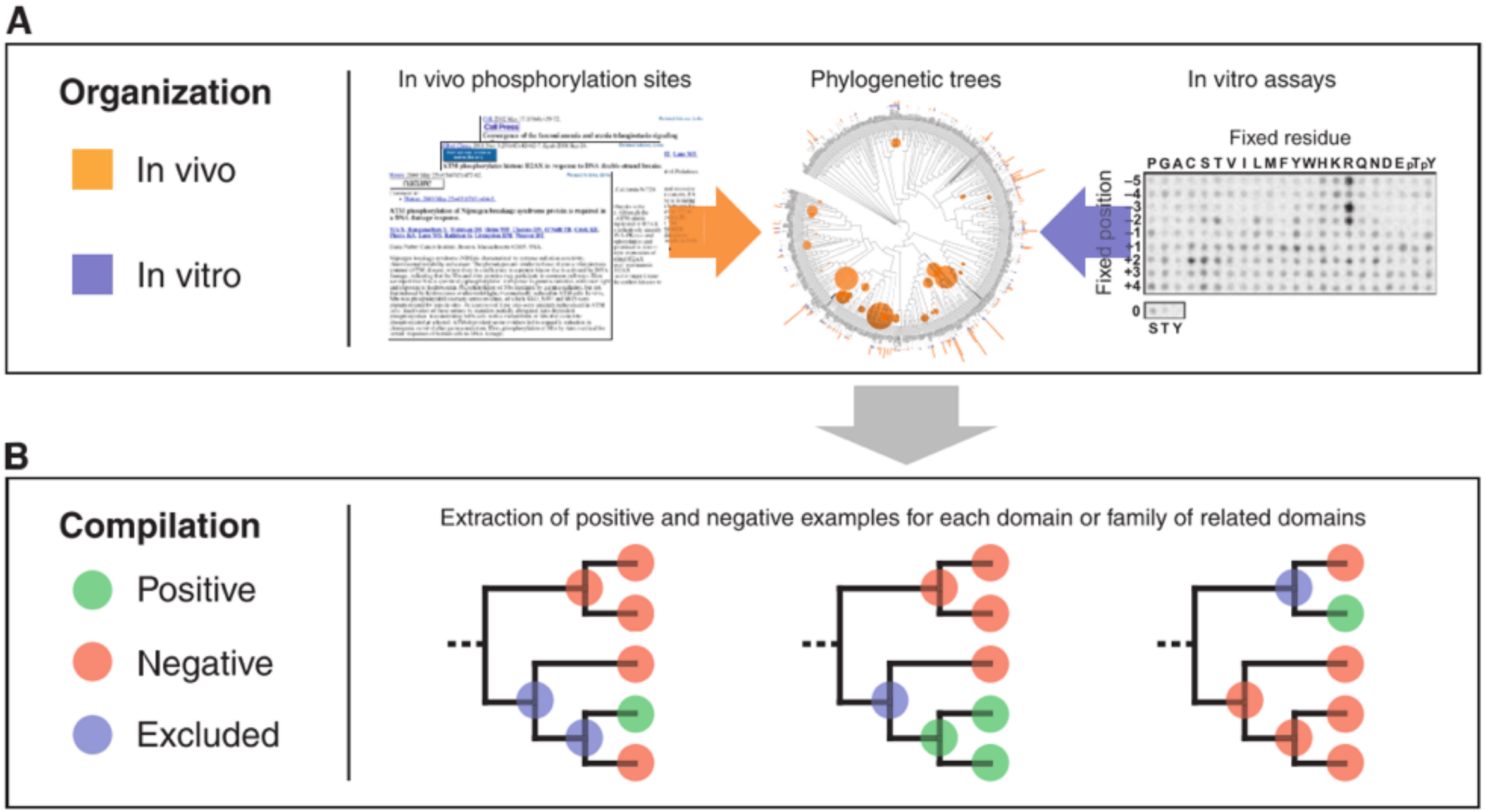
Science Signaling
AAAS

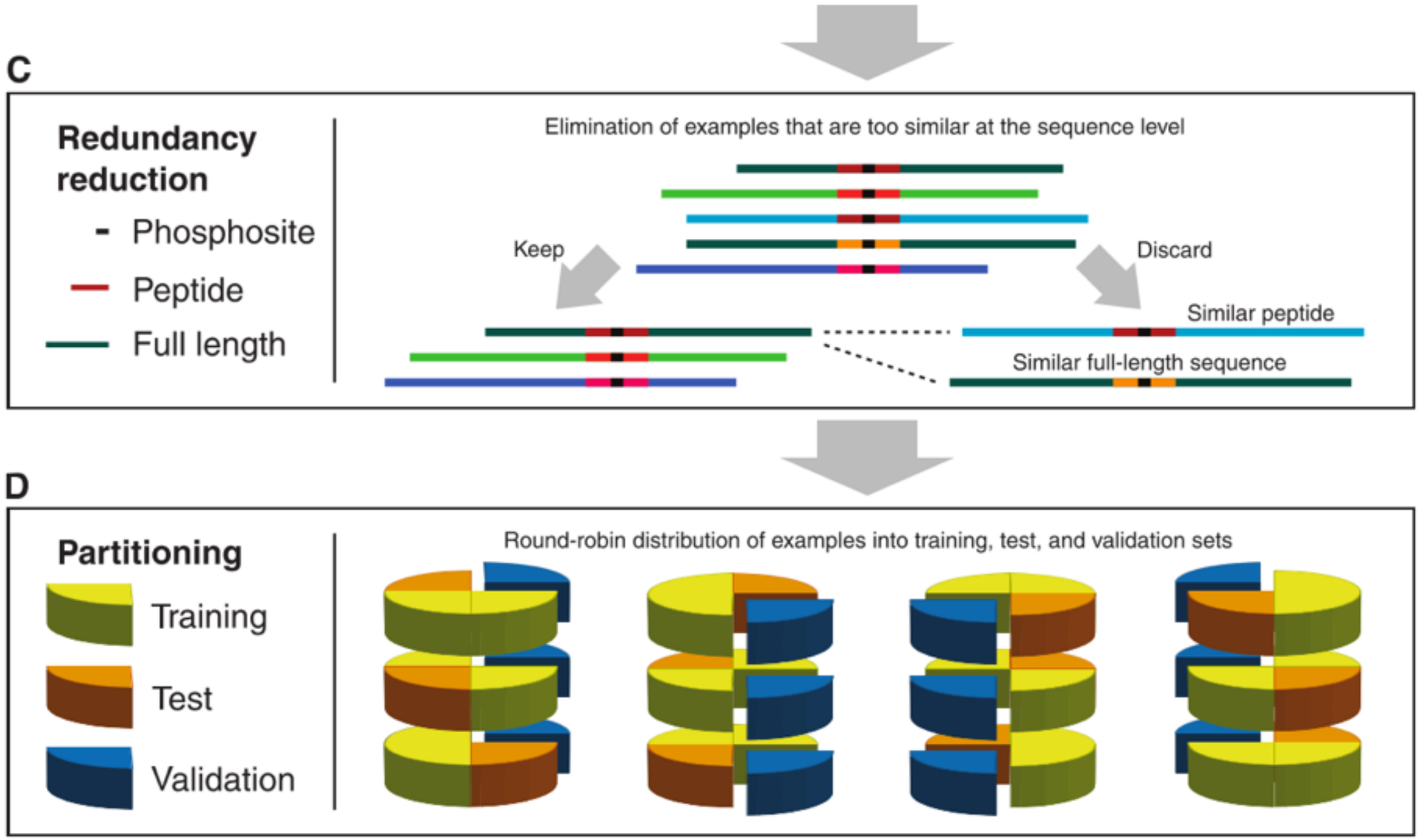
Linear Motif Atlas for Phosphorylation-Dependent Signaling

Martin Lee Miller,[1,2*] Lars Juhl Jensen,[2,3*] Francesca Diella,[3] Claus Jørgensen,[4] Michele Tinti,[5] Lei Li,[6] Marilyn Hsiung,[4] Sirlester A. Parker,[7] Jennifer Bordeaux,[7] Thomas Sicheritz-Ponten,[1] Marina Olhovsky,[4] Adrian Pasculescu,[4] Jes Alexander,[8] Stefan Knapp,[9] Nikolaj Blom,[1] Peer Bork,[2,10] Shawn Li,[6] Gianni Cesareni,[5] Tony Pawson,[4] Benjamin E. Turk,[7] Michael B. Yaffe,[8†] Søren Brunak,[1,2†] Rune Linding[4,8,11†]

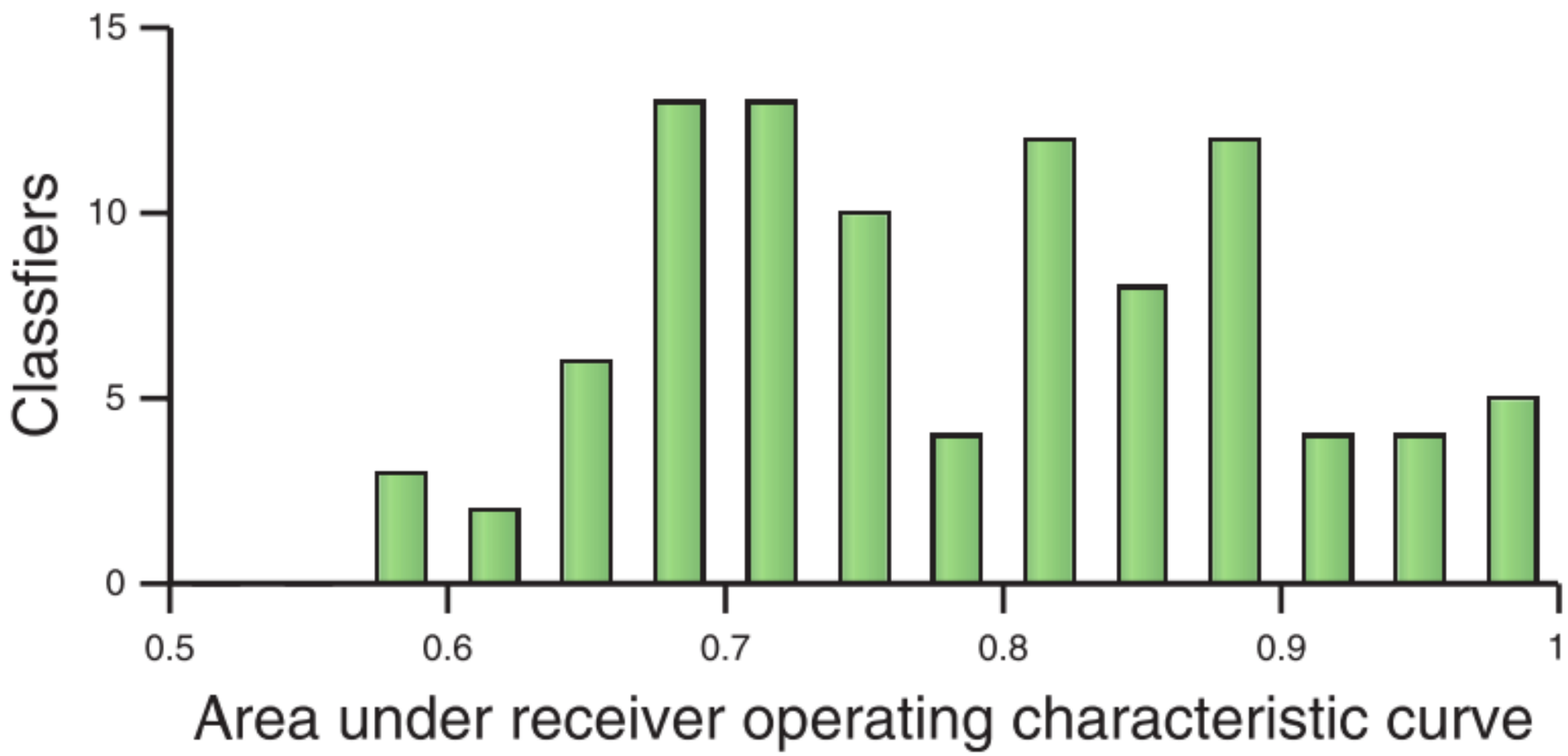
(Published 2 September 2008)


Protein domain sequences
Classified phosphorylation sites
In vitro specificity assays
Pipeline
DNAPK
SMG1
TRRAP
mTOR
ATR
ATM
Motif atlas
CK2 family
Bits
3.0
1.5
0.0
N
C
Design of consensus antibodies
XXXpSDED
XXXpSEED
XXXpSDDE
XXXpTDED
Modeling of signaling networks
Rad50
DNA-PK
ATM
CK2A2
Detection of purification biases
PAC
IMAC
TiO2


A
Organization
In vivo
In vitro
In vivo phosphorylation sites
Phylogenetic trees
In vitro assays
Fixed residue
P G A C S T V I L M F Y W H K R Q N D E pT pY
Fixed position
-5
-4
-3
-2
-1
+1
+2
+3
+4
0
S T Y
B
Compilation
Positive
Negative
Excluded
Extraction of positive and negative examples for each domain or family of related domains


C
Redundancy reduction
Phosphosite
Peptide
Full length
Elimination of examples that are too similar at the sequence level
Keep
Discard
Similar peptide
Similar full-length sequence
D
Partitioning
Training
Test
Validation
Round-robin distribution of examples into training, test, and validation sets

Overview of the performance of the NetPhorest classifiers



NetworKIN algorithm

Resource

Cell

Systematic Discovery of In Vivo Phosphorylation Networks

Rune Linding,[1,2,7,*] Lars Juhl Jensen,[3,7] Gerard J. Ostheimer,[2,4,7] Marcel A.T.M. van Vugt,[2,5] Claus Jørgensen,[1] Ioana M. Miron,[1] Francesca Diella,[3] Karen Colwill,[1] Lorne Taylor,[1] Kelly Elder,[1] Pavel Metalnikov,[1] Vivian Nguyen,[1] Adrian Pasculescu,[1] Jing Jin,[1] Jin Gyoon Park,[1] Leona D. Samson,[4] James R. Woodgett,[1] Robert B. Russell,[3] Peer Bork,[3,6,*] Michael B. Yaffe,[2,*] and Tony Pawson[1,*]

2007

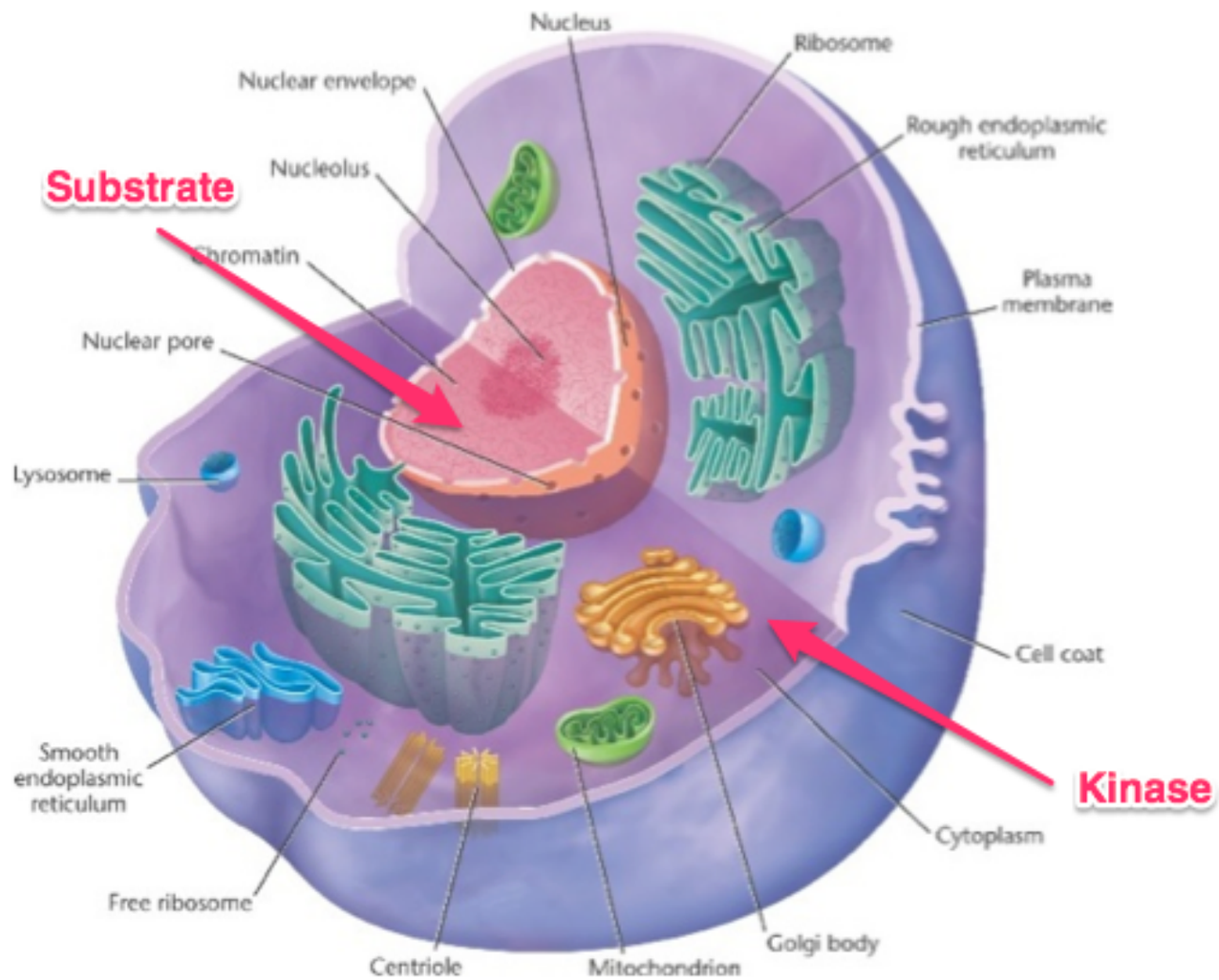
- Sequence motifs lack sufficient information to uniquely identify the physiological substrates (Linding et al. 2007)

- Cellular context:
  - Subcellular compartmentalization
  - Colocalization via anchoring protein & scaffolds
  - Substrate capture by noncatalytic interaction domain
  - Temporal & cell type specific coexpression
  - Kinase-docking motifs within substrates
  - Regulatory subunits

Cellular context - Subcellular compartmentalization

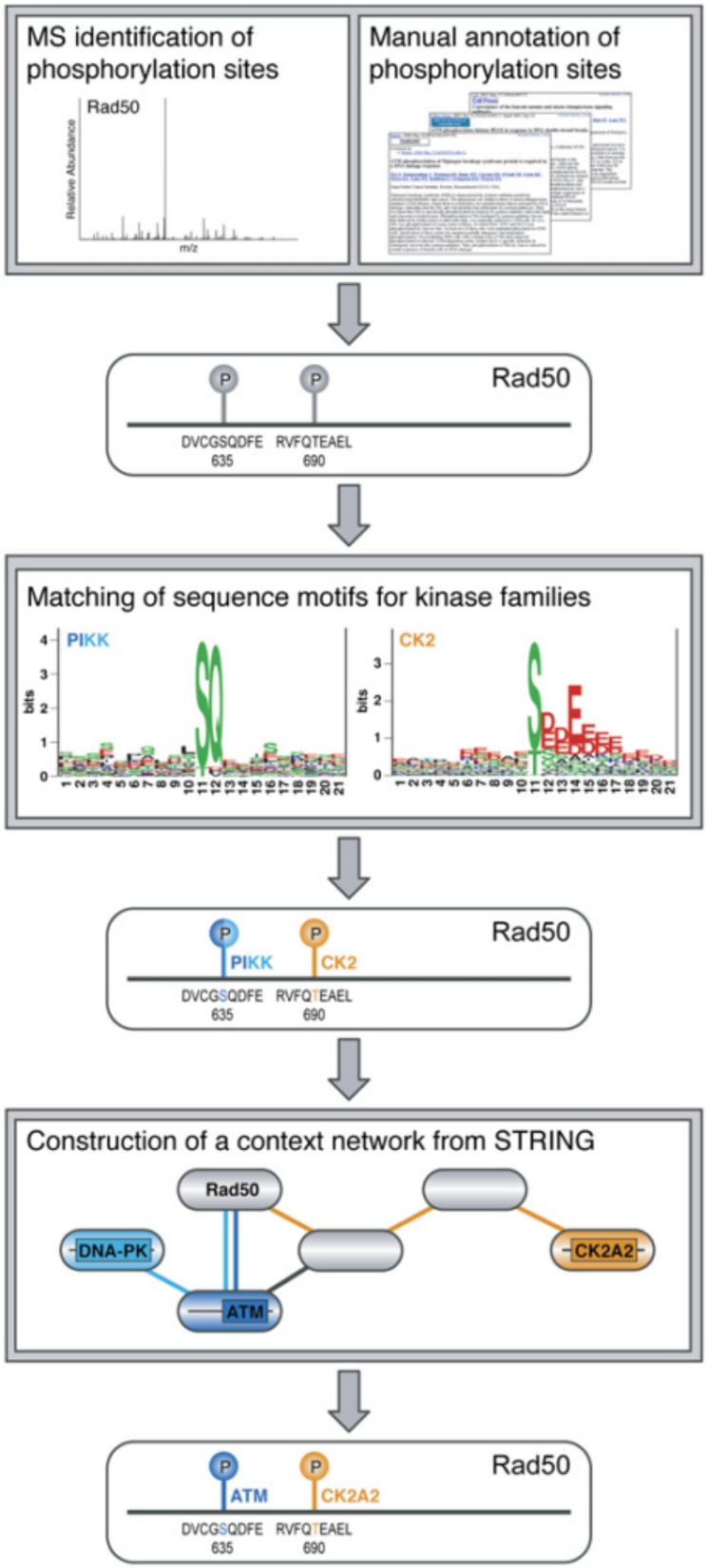
A kinase and a substrate should exist in the same place at the same time.

eg) Subcellular compartmentalization

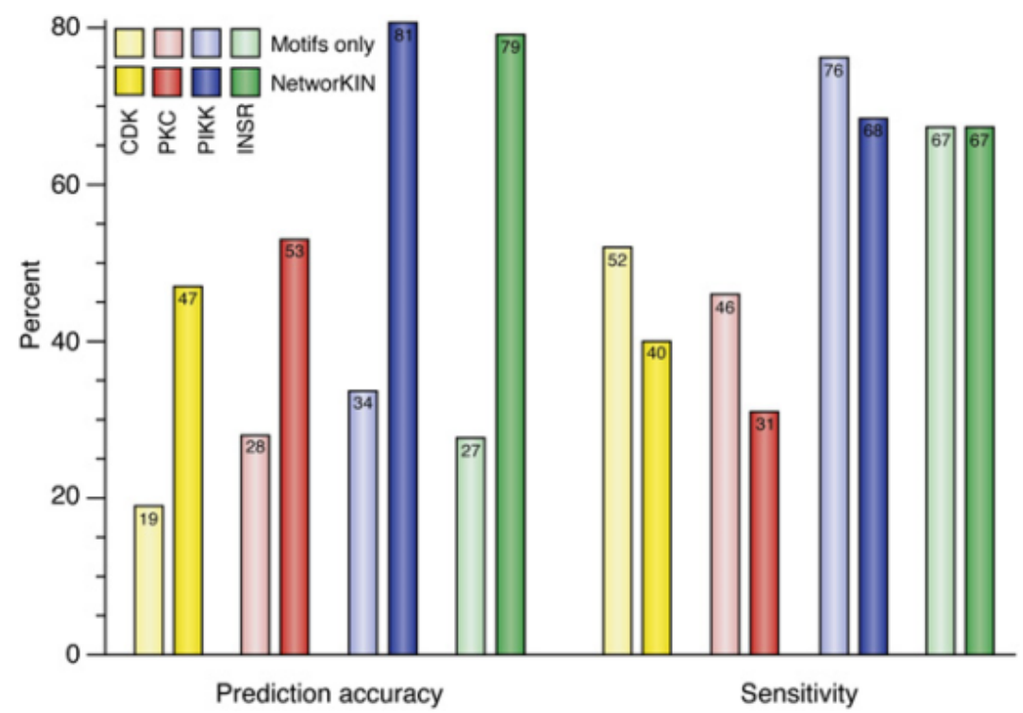


http://bio3400.nicerweb.com

NetworKIN Algorithm



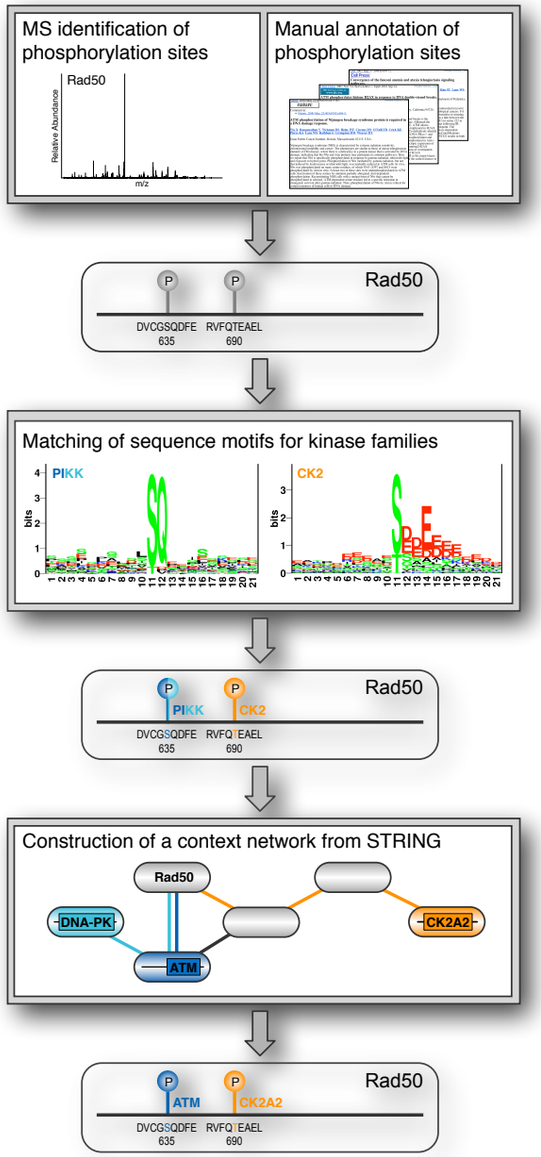
- Scansite
- NetPhosK



Effects of Including Substrate Context

New NetworKIN Framework

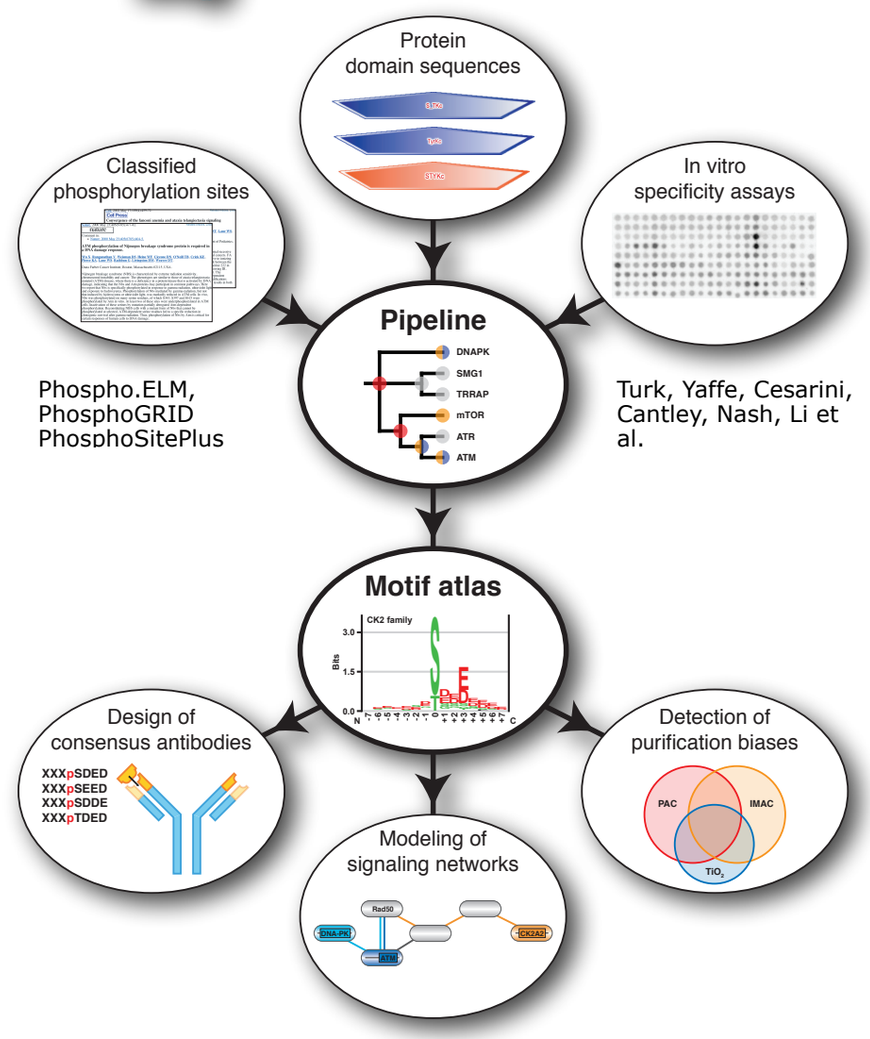
NetworKIN



Linding et al. Cell, 129,

http://NetworKIN.info

NetPhorest



Miller et al. Science Signaling 2008 Sep 2;1(35):ra2.

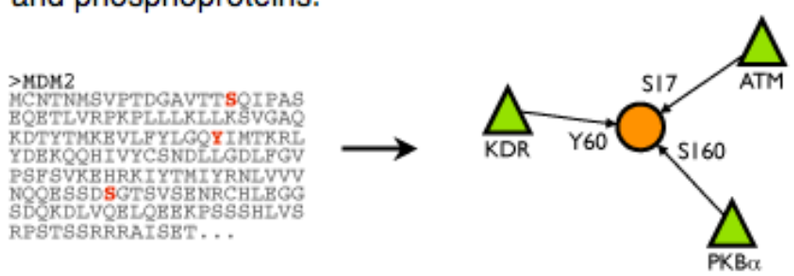
http://NetPhorest.info

KinomeXplorer

KinomeXplorer is an integrated framework for modeling kinase-substrate interactions and aid in the design of inhibitor-based follow-up perturbation experiments. An interactive web interface allows investigation of predicted kinase-substrate interactions from human and major eukaryotic model organisms.

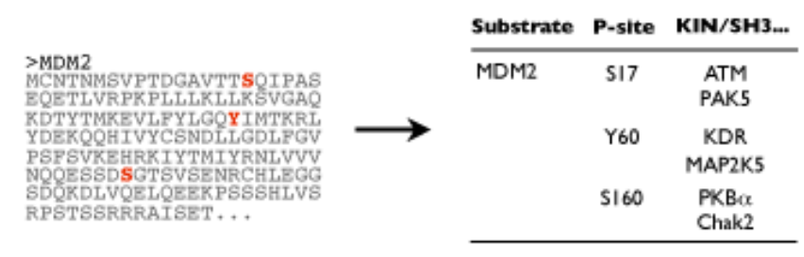
NetworKIN

NetworKIN is a method for predicting in vivo kinase-substrate relationships, that augments intrinsic specificities of kinases with cellular context for kinases and phosphoproteins.



NetPhorest

NetPhorest is a non-redundant collection of 222 sequence-based classifiers for linear motifs in phosphorylation-dependent signaling, including kinases, phosphatases, SH3 domains and more.



Additional resource: KinomeSelector which enables the user select an optimal kinase panel to assess the kinome-wide selectivity of kinase inhibitors.

DTU Technical University of Denmark

CENTERFOR BIOLOGICAL SEQUENCE ANALYSIS CBS

Cellular Signal Integration Group

The Novo Nordisk Foundation Center for Protein Research

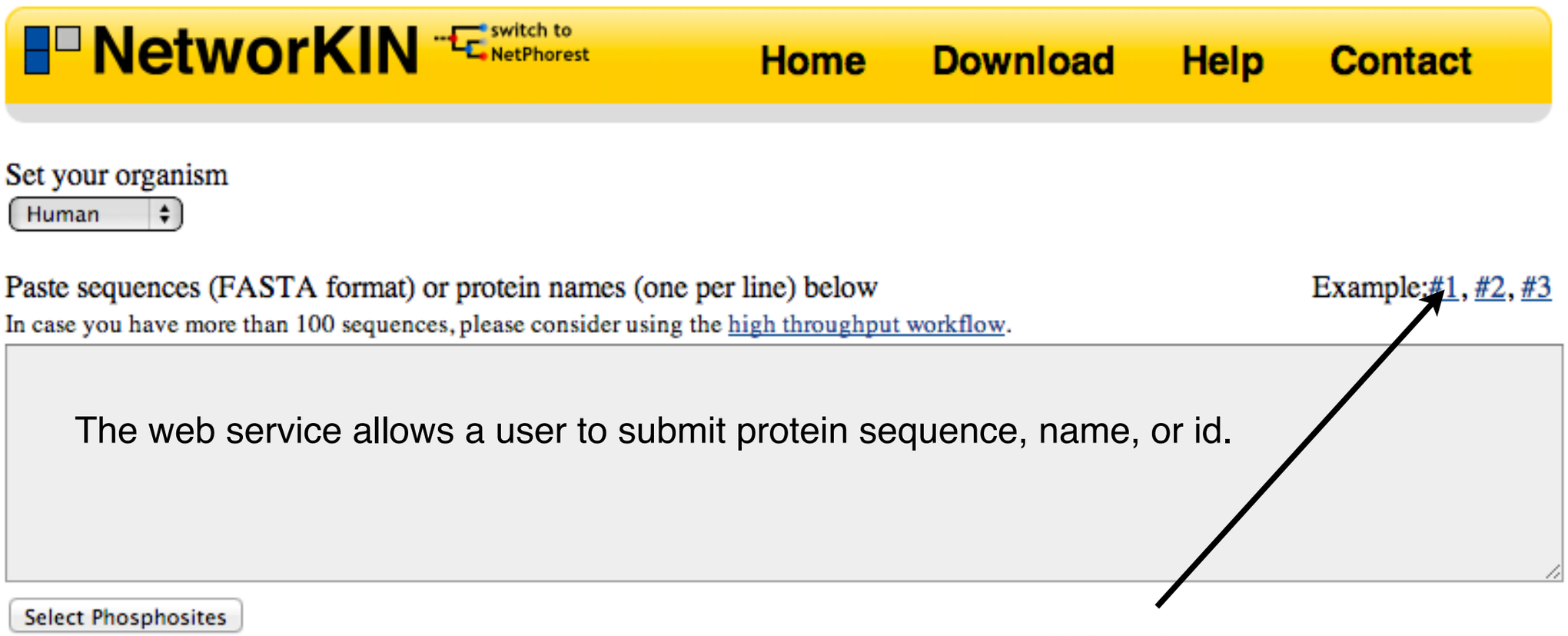
(C) 2005-2013 Rune Linding and Lars Juhl Jensen

Web-development: Heiko Horn and Jinho Kim

GETFIREFOX W3C CSS 2 W3C XHTML 1.0 POSTGRESQL POWERED APACHE python

http://kinomexplorer.info

NetworKIN is an integrated framework for modeling kinase-substrate interactions and aid in the design of inhibitor-based follow-up perturbation experiments. An interactive web interface allows investigation of predicted kinase-substrate interactions from human and major eukaryotic model organisms.



The web service allows a user to submit protein sequence, name, or id.

Click

Let's start with example sequences

http://kinomexplorer.info/
http://networkin-beta.cbs.dtu.dk/

NetworKIN switch to NetPhorest

Home Download Help Contact

Set your organism

Human

Paste sequences (FASTA format) or protein names (one per line) below

Example:#1, #2, #3

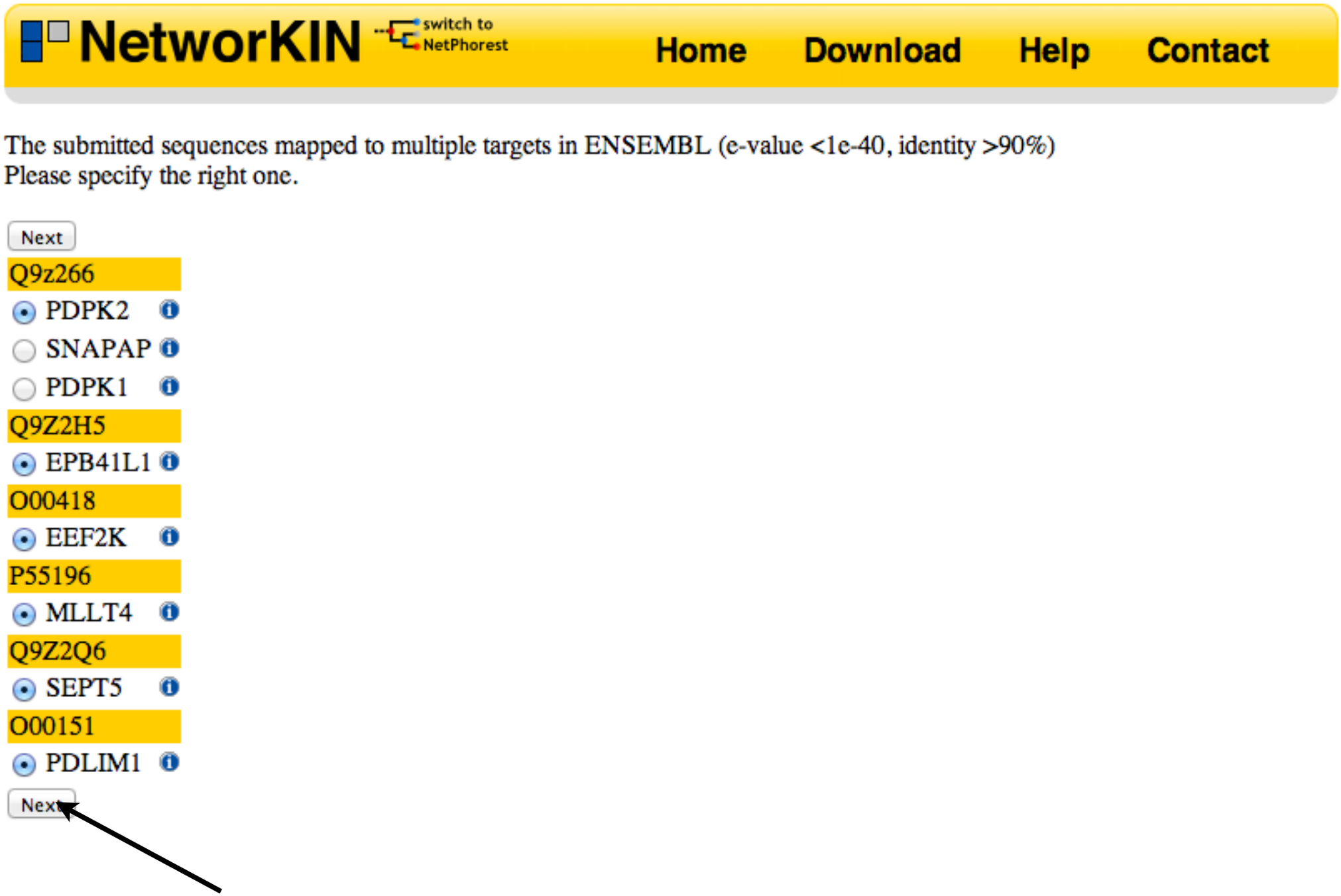
In case you have more than 100 sequences, please consider using the high throughput workflow.

```
>O00151
MTTQQIDLQGPGPWGFRLVGGKDFEQPLAISRVTPGSKAALANLCIGDVITAIDGENTSNMTHLEAQNRIKGCTDNLTLTVARSEHKVWSPLVTEEGKRHPYKMNLASEPQEVL
HIGSAHNRSAMPFTASPASSTTARVITNQYNNPAGLYSSENISNFNNALESKTAASGVEANSRPLDHAQPPSSLVIDKESEVYKMLQEKQELNEPPKQSTSFLVLQEILESEEKGDPN
KPSGFRSVKAPVTKVAASIGNAQKLPMCDKCGTGIVGVFVKLRDRHRHPECYVCTDCGTNLKQKGHFFVEDQIYCEKHARERVTPPEGYEVVTVFPKSQSQSQSQSQSQ
>O00418
MADEDLIFRLEGVDGGQSPRAGHDGDSDGDSDDEEGYFICPITDDPSSNQNVNSKVNKYYSNLTKSERYSSSGSPANSFHFKEAWKHAIQKAKHMPDPWAEFHLEDIATERATR
HRYNAVTGEWLDDEVLIKMASQPFGRGAMRECFRTKKLSNFLHAQQWKGASNYVAKRYIEPVDRDVYFEDVRLQMEAKLWGEEYNRHKPPKQVDIMQMCIIELKDRPGKPLFH
LEHYIEGKYIKYNSNSGFVRDDNIRLTPQAFSHFTFERSGHQLIVVDIQGVGDLYTDPQIHTETGTDFGDGNLGVRGMALFFYSHACNRICESMGLAPFDLSPRERDAVNQNTKLL
QSAKTILRGTEEKCGSPQVRTLSGSRPPLLRPLSENSGDENMSDVTFDSLPSSPSSATPHSQKLDHLHWPVFSDLDNMASRDHDHLDNHRESENSGDSGYPSEKRGELDDPEPRE
HGHSYSNRKYESDEDSLGSSGRVCVEKWNLLNSSRLHLPRASAVALEVQRLNALDLEKKIGKSILGKVHLAMVRYHEGGRFCEKGEEWDQESAVFHLEHAANLGELEAIVGLGLM
YSQLPHHILADVSLKETEENKTKGFDYLLKAAEAGDRQSMILVARAFDSGQNLSPDRCQDWLEALHWYNTALEMTDCDEGGEYDGMQDEPRYMMLAREAEMLFTGGYGLEK
DPQRSGDLYTQAAEAAMEAMKGRLANQYYQKAEEAWAQMEE
>P55196
MSAGGRDEERRKLADIIHHWNANRLDLFEISQPTEDLEFHGVMRFYFQDKAAGNFATKCIRVSSTATTQDVIETLAEKFRPDMRMLSSPKYSLYEVHVSGERRLDIDEKPLVVQLN
WNKDDREGRFVLKNENDAIPPKAQSNGPEKQEKEGVIQNFKRTLSKKEKKEKKKREKEALRQASDKDDRPFQGEDVENSRLAAEVYKDMPETSFTRTISNPEVVMKRRRQQKLE
KRMQEFRSSDGRPDSGGTLRIYADSLKPNIPYKTILLSTTDPADFAVAEALEKYGLEKENPKDYCIARVMLPPGAQHSDEKGAKEIILDDDECPLQIFREWPSDKGILVFQLKRRPPD
HIPKKTKKHLEGKTPKGKERADGSGYGSTLPPEKLPYLVELSPDGSDSRDKPKLYRLQLSVTEVGTEKLDDNSIQLFGPGIQPHHCDLTNMDGVVTVTPRSMDAETYVEGQRISET
TMLQSGMKVQFGASHVFKFVDPSQDHALAKRSVDGGLMVKGPRHKPGIVQETTFDLGGDIHSGTALPTSKSTTRLDSDRVSSASSTAERGMVKPMIRVEQQPDYRRQESRTQD
ASGPELILPASIEFRESSEDSFLSAIINYTNSSTVHFKLSPTYVLYMACRYVLSNQYRPDISPTERTHKVIAVVNKMVSMMEGVIQKQKNIAGALAFWMANASELLNFIKQDRDLSRIT
LDAQDVLAHLVQMAFKYLVHCLQSELNNYMPAFLDDPEENSLQRPKIDDVLHTLTGAMSLLRRCRVNAALTIQLFSQLFHFINMWLFNRLVTDPDSGLCSHYWGAIIRQQLGHI
```

Select Phosphosites

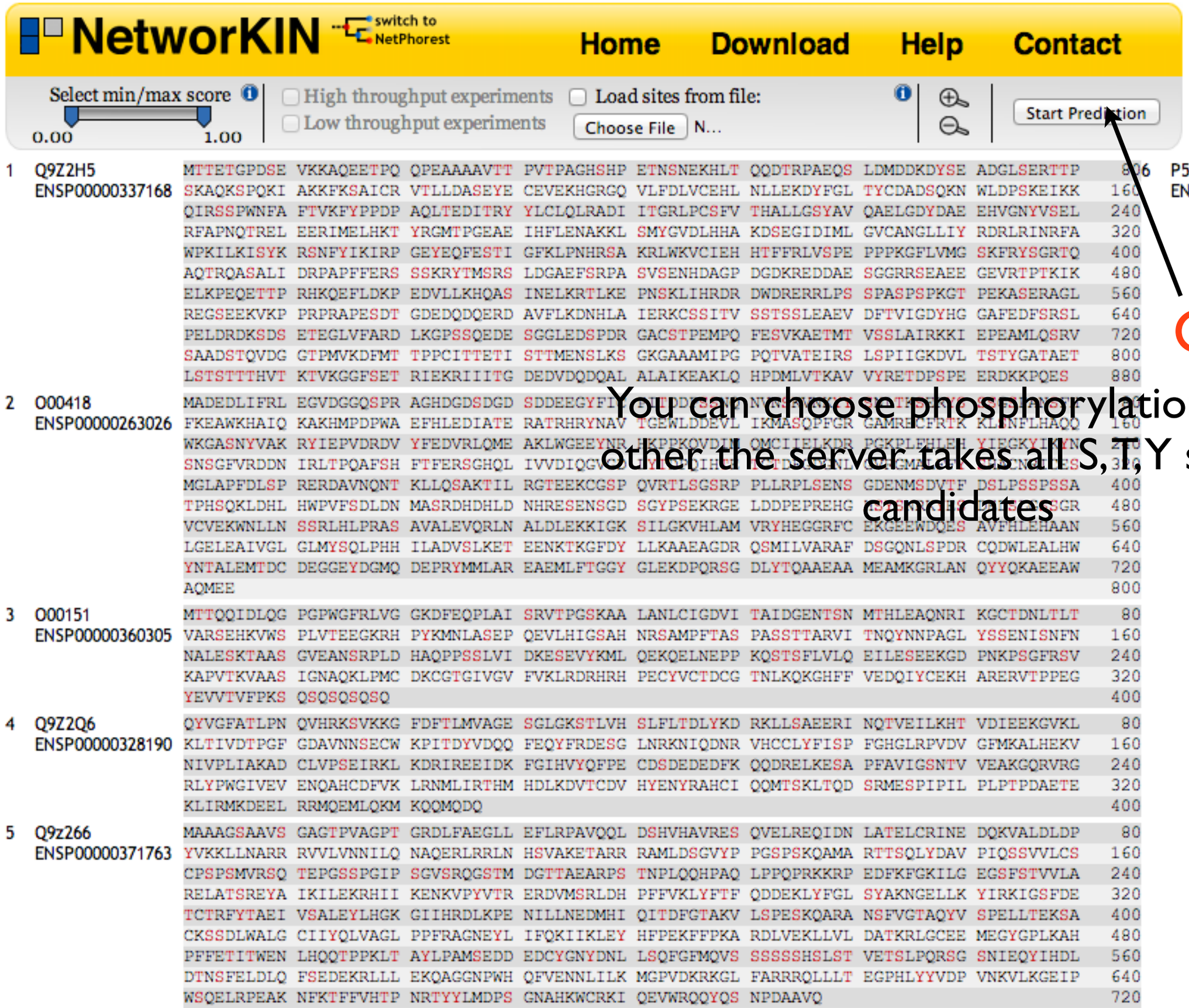
Click

NetworKIN suggests proteins based on sequence homology



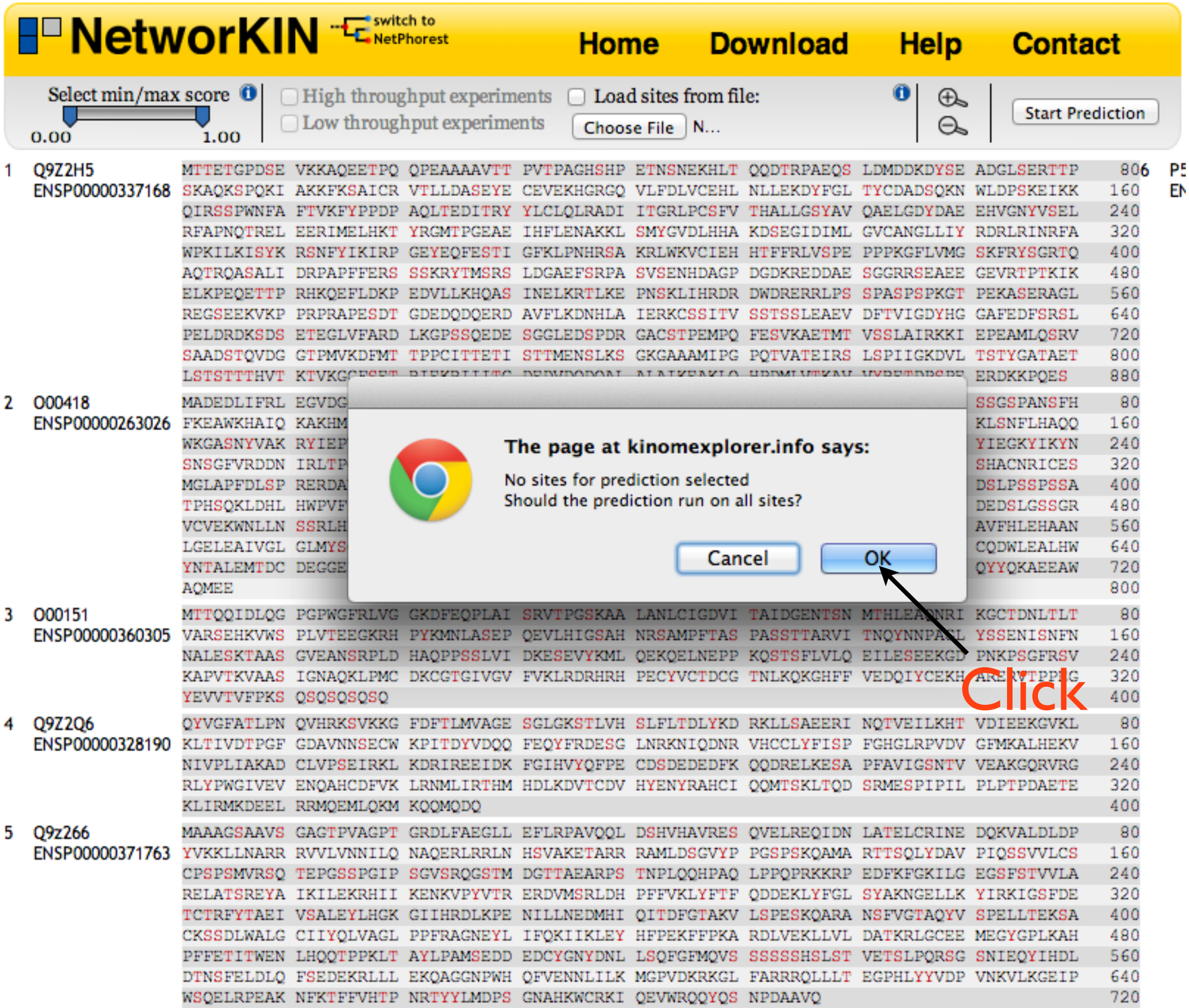
Click

after confirming that the suggested proteins are right ones



Click

You can choose phosphorylation sites, other the server takes all S, T, Y sites as candidates


NetworKIN
switch to NetPhorest
Home
Download
Help
Contact
Select min/max score
0.00
1.00
High throughput experiments
Low throughput experiments
Load sites from file:
Choose File
N...
Start Prediction
1 Q9Z2H5 ENSP00000337168
2 O00418 ENSP00000263026
3 O00151 ENSP00000360305
4 Q9Z2Q6 ENSP00000328190
5 Q9z266 ENSP00000371763
The page at kinomexplorer.info says:
No sites for prediction selected
Should the prediction run on all sites?
Cancel
OK
Click

Results

NetworKIN switch to NetPhorest

Home Download Help Contact

Minimum score 2.00 | Max. difference 1.00 | ☑KIN ☐14-3-3 ☐BRCT ☑PTB ☑SH2 ☐WW | Max. # of Predictions 2 Update | Real time filter for results ☑ | save

000151

000418

S18	Kinase	MAPK13	2.0509	VDGGQSPRAGH
S74	Kinase	MAPK13	2.0509	YSSSGSPANSF
T254	Kinase	MAPK13	2.0168	DNIRLTPQAFS
S359	Kinase	MAPK13	2.0168	EEKCGSPQVRT
S366	Kinase	p70S6Kb	2.8915	QVRTLSGSRPP
		p70S6K	2.6063	QVRTLSGSRPP
S396	Kinase	MAPK13	2.0509	DSLPSSPSSAT
T401	Kinase	MAPK13	2.0509	SPSSATPHSQK
S464	Kinase	p70S6Kb	2.5262	HGHSYSNRKYE
		p70S6K	2.3401	HGHSYSNRKYE

P55196

Y1235	PTB	FRS2	2.0402	NQWPNYEEKPH

Q9Z2H5

Q9Z2Q6

Q9z266

T20	Kinase	PKCiota	2.0760	PVAGPTGRDLF
S112	Kinase	PKCiota	2.0760	RRLNHSVAKET
T117	Kinase	PKCiota	2.2127	SVAKETARRRA
S301	Kinase	PKCiota	2.3665	LYFGLSYAKNG
T367	Kinase	PKCiota	2.2127	ITDFGTAKVLS

Kinases that are predicted to phosphorylate the site

Residues that are predicted to be phosphorylated

Exercise 1.

Q. Using NetworKIN, find top 3 kinases that are predicted to phosphorylate S39 position in human N-RAS protein.

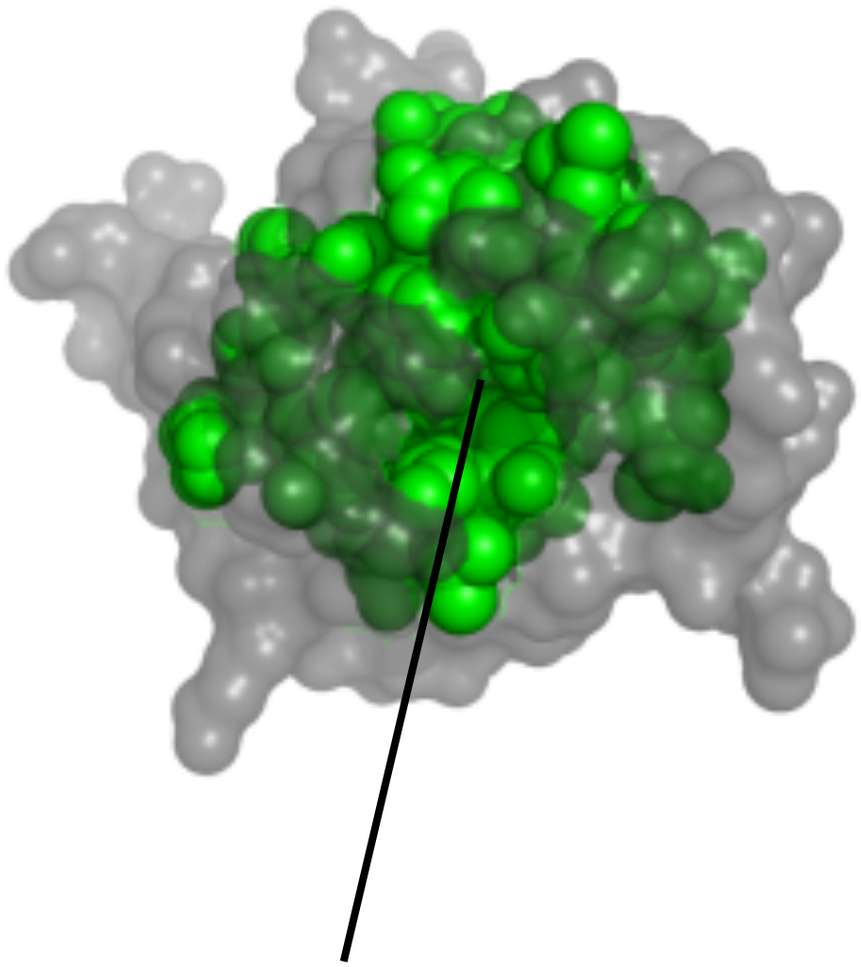
Protein sequence of human N-RAS

```
>sp|P01111|RASN_HUMAN GTPase NRas OS=Homo sapiens GN=NRAS PE=1 SV=1
MTEYKLVVVGAGGVGKSALTIQLIQNHFVDEYDPTIEDSYRKQVVIDGETCLLDILDTAG
QEEYSAMRDQYMRTGEGFLCVFAINNSKSFADINLYREQIKRVKDSDDVPMVLVGNKCDL
PTRTVDTKQAHELAKSYGIPFIETSAKTRQGVEDAFYTLVREIRQYRMKKLNSSDDGTQG
CMGLPCVVM
```

NetworKIN web site

http://kinomexplorer.info/
http://networkin-beta.cbs.dtu.dk/

Not all the residues are accessible by kinases.



Core residues are not accessible by kinases

http://pipe.scs.fsu.edu/

Exercise2. Predict phosphosites of MDM2 (human) and kinases that phosphorylate them.

- Find the sequence at Uniprot.org.
- Use NetPhos score cutoff 0.7 for phosphosite prediction
- Use NetworKIN score cutoff 20 for kinase prediciton

How many phosphosite-kinase interactions were predicted?
Which one is the most probable one?

Exercise3. Add S395 manually and run again with the same cutoffs.

How many phosphosite-kinase interactions were predicted?
Which one is the most probable one?

High-throughput workflow of NetworKIN

Sample

Thermo Scientific Orbitrap Fusion Tribrid LC-MS (r) with Thermo Scientific Dionex UltiMate 3000 Series UHPLC.

Thousands of phosphosites
on thousands of proteins

You are only interested in the detected phosphosites, not all S, T, Y sites

NetworKIN switch to NetPhorest

Home Download Help About

Set your organism

Human

Paste sequences (FASTA format) or protein names (one per line) below

Example:#1, #2, #3

In case you have more than 100 sequences, please consider using the high throughput workflow.

Select Phosphosites

Click

NetworKIN switch to NetPhorest

Home Download Help About

This is the high-throughput interface for NetworKIN. Click here to go back to the low-throughput version.

Select the sequence database:

Ensembl 74 (human)

Put modified sites below in the format Protein ID, Position, Residue (space/tab delimited) Example: #1, #2, ProteomeDiscoverer

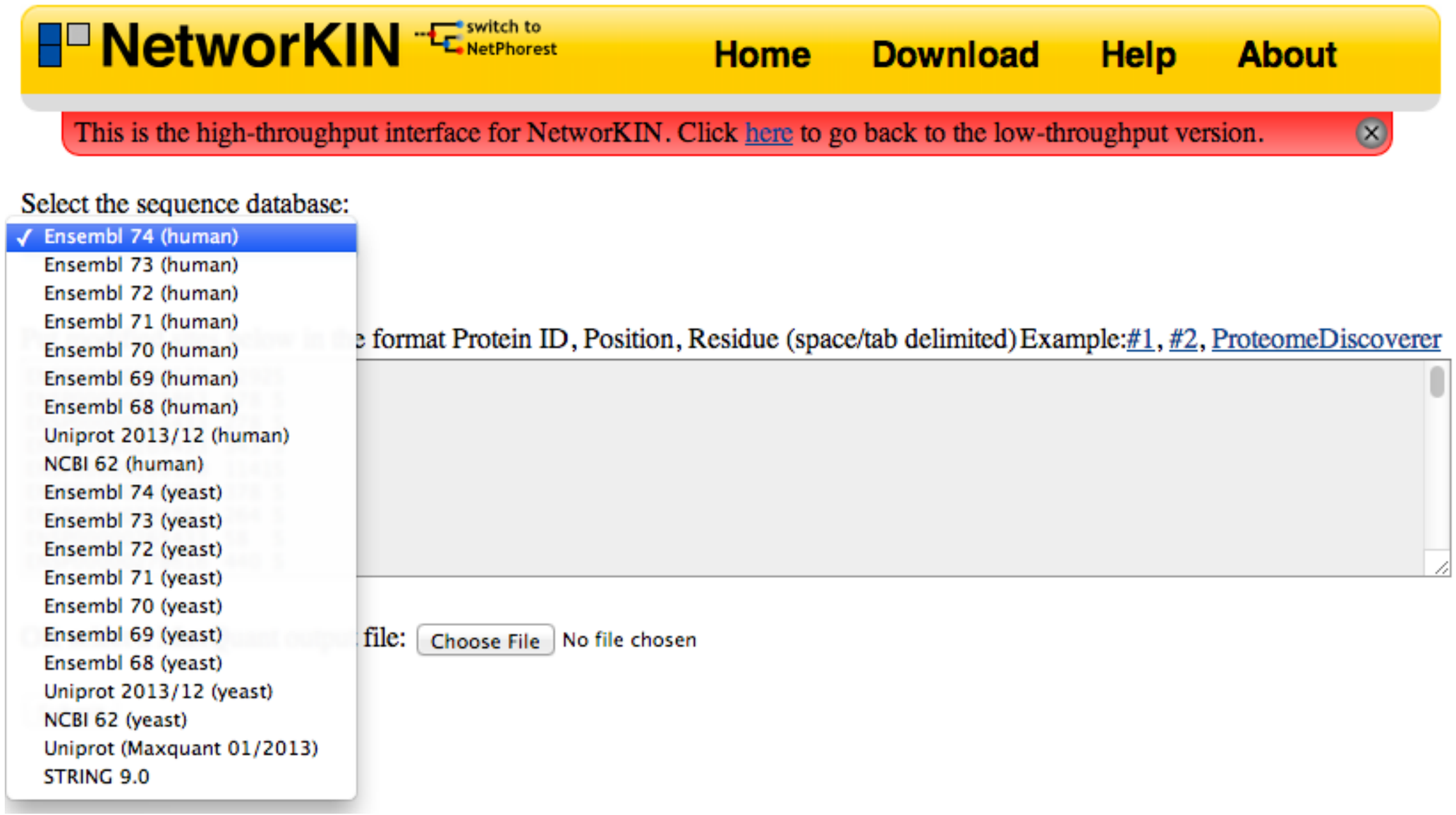
```
ENSP00000298139  1292S
ENSP00000325863  678  S
ENSP00000265433  278  S
ENSP00000265433  343  S
ENSP00000298139  1141S
ENSP00000269305  378  S
ENSP00000325863  264  S
ENSP00000265433  58   S
ENSP00000278616  440  S
```

OR select a MaxQuant output file: Choose File No file chosen

Submit

Raw output from the search engine

Choose the right sequence database and version



NetworKIN switch to NetPhorest

Home Download Help About

Not all submitted identifier could not be mapped to a unique STRING entity.
Please select a desired isotype or homolog.

Next

ENSP00000355746

- (•) PSEN2 (100.00% 448AA)
- ( ) PSEN1 (64.81% 425AA)

ENSP00000352925

- (•) TRRAP (99.22% 3859AA)
- ( ) ATR (24.02% 226AA)

ENSP00000402030

- (•) LIG4

ENSP00000366863

- (•) TBC1D4

ENSP00000269305

- (•) TP53

ENSP00000298139

- (•) WRN

ENSP00000347232

- (•) BLM

ENSP00000278616

- (•) ATM

ENSP00000295266

- (•) PDHA2

ENSP00000358576

- (•) DCLRE1B

NetworKIN switch to NetPhorest

Home Download Help About

Not all submitted identifier could not be mapped to a unique STRING entity.
Please select a desired isotype or homolog.

Click

Click

Next

ENSP00000355746

- PSEN2 (100.00% 448AA)
- PSEN1 (64.81% 425AA)

ENSP00000352925

- TRRAP (99.22% 3859AA)
- ATR (24.02% 226AA)

ENSP00000402030

- LIG4

ENSP00000366863

- TBC1D4

ENSP00000269305

- TP53

ENSP00000298139

- WRN

ENSP00000347232

- BLM

ENSP00000278616

- ATM

ENSP00000295266

- PDHA2

ENSP00000358576

- DCLRE1B

Reflect - PSEN1

Protein Add About

PSEN1 (ENSP00000326366) H. sapiens Edit

PS1; presenilin-1; presenilin 1; presenilin1; PSEN-1; C-PS1

PSN1_HUMAN, Sequence, Domains, Structure, Locus, Literature

MTELPAPLSYFQNAQMSEDNHLSNTVRSQNDNRERQEHNDRRSLGHPI

No structure information available

Presenilin-1 (EC 3.4.23.-) (PS-1) (Protein S182) [Contains: Presenilin-1 NTF subunit; Presenilin-1 CTF subunit; Presenilin-1 CTF12 (PS1-

NetworKIN switch to NetPhorest

Home Download Help About

Minimum score 2.00 | Max. difference 4.00 | ☑KIN ☐SH2 ☐PTP ☐PTB ☑14-3-3 ☑BRCT ☐WW ☑WD40 | Max. # of Predictions 10 Update | Real time filter for results ☑ | save

ENSP00000219476

S1341	Kinase	PDHK1	8.24	SSSVSSQEEKS
		GSK3beta	5.77	SSSVSSQEEKS

ENSP00000261716

S554	Kinase	ATM	128.36	NSFLESQKREY
S990	Kinase	ATM	95.10	STSVTSQISNG

ENSP00000263710

S1196	Kinase	CK1delta	8.02	KFSFRSQEDLN
		CK1epsilon	5.16	KFSFRSQEDLN

ENSP00000265335

S635	Kinase	ATM	167.80	FDVCGSQDFES
	BRCT	BRCA1	3.98	FDVCGSQDFES

ENSP00000265433

S58	Kinase	ATM	228.14	SVTNLSQTDEI
S278	Kinase	ATM	228.14	TGITNSQTLIP
	BRCT	BRCA1	3.74	TGITNSQTLIP
		MDC1	3.42	TGITNSQTLIP
S343	Kinase	ATM	228.14	PGPSLSQGVSV
	BRCT	BRCA1	3.74	PGPSLSQGVSV
		MDC1	3.42	PGPSLSQGVSV
S397	Kinase	ATM	214.52	KFRMLSQDAPT
S615	Kinase	ATM	213.95	ESSKISQENEI

Change

NetworKIN switch to NetPhorest

Home Download Help About

Minimum score 20.00 | Max. difference 4.00 | KIN SH2 PTP PTB 14-3-3 BRCT WW WD40 | Max. # of Predictions 10 Update | Real time filter for results | save

Click

ENSP00000219476

ENSP00000261716

S554	Kinase ATM	128.36	NSFLESQKREY
S990	Kinase ATM	95.10	STSVTSQISNG

ENSP00000263710

ENSP00000265335

ENSP00000265433

S58	Kinase ATM	228.14	SVTNLSQTDEI
S397	Kinase ATM	214.52	KFRMLSQDAPT
S615	Kinase ATM	213.95	ESSKISQENEI

ENSP00000269305

S99	Kinase ATM	167.80	SSSVPSQKTYQ
S378	Kinase ATM	228.14	KGQSTSRHKKL

ENSP00000278616

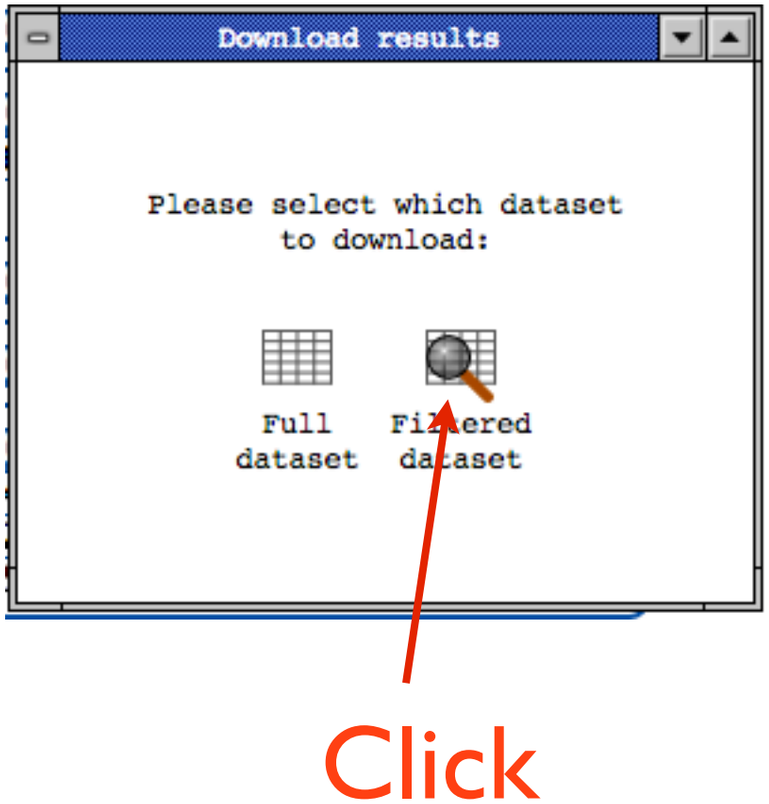
T373	Kinase ATM	151.88	QSYTTTQRESS

ENSP00000288986

ENSP00000295266

ENSP00000298139

S1058	Kinase ATM	91.81	KANTESQSLIL
S1141	Kinase ATM	130.14	KAYSSSQPVIS
S1292	Kinase ATM	130.14	IGMHLSQAVKA


Download results
Please select which dataset
to download:
Full
dataset
Filtered
dataset
Click

Exercise 4. Predict ATM kinase targets with example #1 through high-throughput workflow. (Use a cutoff: 20)

How many phosphosite-ATM kinase interactions were predicted?
Which one is the most probable one?

Exercise 5. Predict phospho-interactions using known phosphosites (phosphosites.txt) through high-throughput workflow. (Choose Ensembl 70 (human) and Use a cutoff: 20)

- How many phosphosite-ATM kinase interactions were predicted?
- Which one is the most probable one?
- Are the substrate and kinase involved in the same biological process?
- What happens if you choose Ensembl 68 (human) instead of Ensembl 70

Exercise 6. Predict kinases that phosphorylate T1162 and T2446 of human MTOR protein through both Low- and High-throughput workflows.